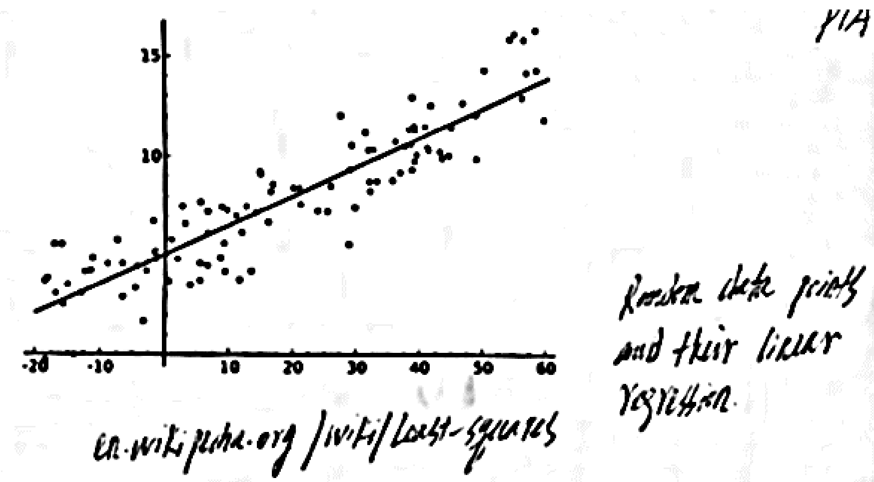
Part I: Least Square Analysis

Case I: Curve Fitting



Case 2: $A_{m\cdot n}\vec{x}_{n\cdot 1} = \vec{b}_{m\cdot 1}$

$m > n$ (when the matrix has more equations than unknown, the matrix A is a tall matrix – so there is usually no solution):

$$\begin{bmatrix} \\ A \\ \\ \end{bmatrix} [\vec{x}] = \begin{bmatrix} \\ \vec{b} \\ \\ \end{bmatrix}$$

Then the error vector can be written as:

$$\vec{e} = \vec{b} - A\vec{x}$$

As it turns out, both cases have the same solution method.

1. Linear Regression

Given data pairs:

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n), n > 2$$

Find the best fit line:

$$y = a_o + a_1 x + e$$

We try to find a way to define the 'best fit' - we can use the length of the error vector itself, defined as an object function.

For each pair (x_i, y_i), the error
$e_i = yi - a_o - a_1 x_i, \quad i = 1, 2, \dots, n$
a_0, a_1; constants to be determined

Object function

$$S_r = e_1^2 + e_2^2 + e_3^2 + \cdots + e_n^2 = \sum_{i=1}^{n} e_i^2 = \sum e_i^2$$

Or

$$S_r = \sum (y_i - a_o - a_1 x_i)^2$$

Minimizing S_r

$$\frac{\delta S_r}{\delta a_o} = 0, \qquad \frac{\delta S_r}{\delta a_1} = 0$$

$$\frac{\delta S_r}{\delta a_o} = \sum 2(y_i - a_0 - a_1 x) \cdot (-1)$$

$$\frac{\delta S_r}{\delta a_o} = (-2) \sum (y_i - a_0 - a_1 x)$$

$$\frac{\delta S_r}{\delta a_1} = \sum 2(y_i - a_0 - a_1 x)(-x_i)$$

$$\frac{\delta S_r}{\delta a_1} = (-2) \sum (y_i - a_0 - a_1 x) \cdot x_i$$

Substituting into the original equation:

$$\sum (y_i - a_0 - a_1 x_i) = 0$$

$$\sum y_i - \sum a_0 - \sum a_1 x_i = 0$$

And

$$\sum (y_i - a_0 - a_1 x_i) x_i = 0$$

$$\sum x_i y_i - \sum a_0 x_i - \sum a_1 x_i^2 = 0$$

$$\rightarrow \quad n \cdot a_0 + \left(\sum x_i\right) a_1 = \sum y_i$$

$$\left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 = \sum x_i y_i$$

Thus, from Gauss-Jordan elimination:

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2}$$

$$a_0 = \frac{1}{n}\left[\sum y_i - \left(\sum x_i\right) a_1\right] = \bar{y} - \bar{x} \cdot a_1$$

Here

$$\bar{x} = \frac{\sum x_i}{n} \quad ; \quad \bar{y} = \frac{\sum y_i}{n}$$

$$A\vec{x} = \vec{b} - A\vec{x}$$

A is a tall matrix (no unique solution – more unknown than equations, or no solutions at all)

Define error vector

$$\vec{e} = \vec{b} - A\vec{x}$$

We try to find the smallest error vector length using the error vector itself. We use the dot product, or transpose multiplied by itself.

Minimize

$$\begin{aligned} S_r &= \vec{e}^T\vec{e} \\ &= \left(\vec{b} - A\vec{x}\right)^T\left(\vec{b} - A\vec{x}\right) \\ &= \left(\vec{b}^T - \vec{x}^T A^T\right)\left(\vec{b} - A\vec{x}\right) \\ &= \overrightarrow{x^T} A^T A\vec{x} - x^T A^T b - b^T A x + b^T b \end{aligned}$$

Two vectors x and y (It's a scalar so order doesn't matter, so the order can be switched with no issues)

$$x^T y = y^T x$$

$$S_r = x^T A^T A\, x - 2xAb + b^T b$$

***Where $A^T A$ is a symmetric matrix*
Note: This is a typical quadratic equation

S_r is a function of vector $\vec{x}$

$$\vec{x} = (x_1, x_2, \dots, x_n)$$

Minimizing S_r

$$\frac{\delta S_r}{\delta x_1} = 0 \quad \frac{\delta S_r}{\delta x_2} = 0 \quad \dots \quad \frac{\delta S_r}{\delta x_n} = 0$$

Or (another form):

$$\frac{\delta S_r}{\delta \vec{x}} = \begin{Bmatrix} \frac{\delta S_r}{\delta x_1} \\ \frac{\delta S_r}{\delta x_2} \\ \dots \\ \frac{\delta S_r}{\delta x_n} \end{Bmatrix} = 0$$

$$\frac{\delta S_r}{\delta \vec{x}} = 2A^T A\vec{x} - 2a^T b$$

From calculus, we knowthe minimum value of this expression is when it is equal to 0.

Find $\vec{\hat{x}}$ such that

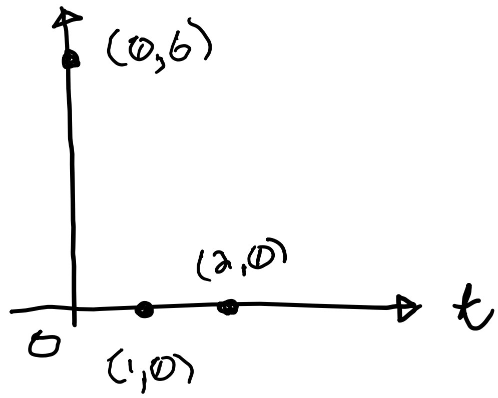
$$A^T A\vec{\hat{x}} = A^T\vec{b}$$

This is how we solve a set (or system of linear equations) when there is no solution.

- We call this the least-squares method
- Essentially, we're just multiplying each side by A^T

Example

Find the closest line to the points (0, 6), (1,0), (2,0)



Solution

Line

$$y = a_0 + a_1 t$$

Point (0,6): $a_0 + a_1(0) = 6$
Point (1,0): $a_0 + a_1(1) = 0$
Point (2,0): $a_0 + a_1(2) = 0$

In matrix form:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} = \begin{Bmatrix} 6 \\ 0 \\ 0 \end{Bmatrix}$$

Convert to:

$$A^T A \vec{x} = A^T \vec{b}$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{Bmatrix} 6 \\ 0 \\ 0 \end{Bmatrix}$$

$$\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} = \begin{Bmatrix} 6 \\ 0 \end{Bmatrix}$$

$$\begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}^{-1} \begin{Bmatrix} 6 \\ 0 \end{Bmatrix} = \begin{Bmatrix} 5 \\ -3 \end{Bmatrix}$$

The line:

$$y = 5 - 3t$$

Geometric explanation

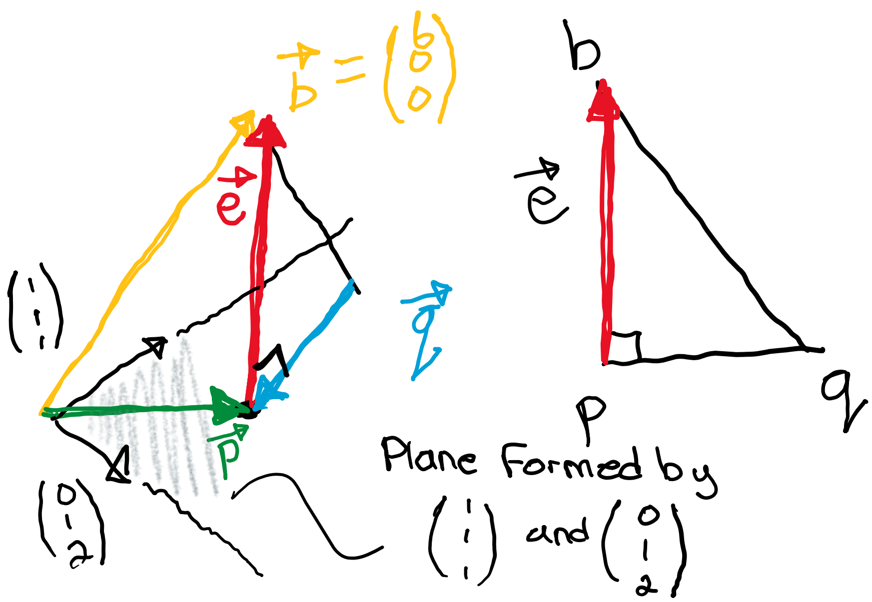
Another way of thinking about the least-squares solution:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} = \begin{Bmatrix} 6 \\ 0 \\ 0 \end{Bmatrix}$$

$$a_0 \begin{Bmatrix} 1 \\ 1 \\ 1 \end{Bmatrix} + a_1 \begin{Bmatrix} 0 \\ 1 \\ 2 \end{Bmatrix} = \begin{Bmatrix} 6 \\ 0 \\ 0 \end{Bmatrix}$$

The column vectors of A: $\begin{Bmatrix} 1 \\ 1 \\ 1 \end{Bmatrix}$ and $\begin{Bmatrix} 0 \\ 1 \\ 2 \end{Bmatrix}$ will expand a plane in 3D (3 dimensions)

$b = \begin{Bmatrix} 6 \\ 0 \\ 0 \end{Bmatrix}$ does not belong to the plane.



$$\vec{b} = \vec{p} + \vec{e}$$

And:

$$A\vec{\hat{x}} = \vec{p}$$

$$\vec{e} = \vec{b} - \vec{p}$$

Is the smallest value when $\vec{p}$ is a projection of $\vec{b}$ onto the plane formed by the columns of matrix A.